DOCUMENT RESUME

ED 324 156 RC 017 747

AUTHOR Williams, Bev; Gross, Kathy
TITLE English Proficiency Test and Classroom Application.
PUB DATE Feb 90
NOTE 19p.; Paper presented at the Annual Alaska Bilingual Multicultural Education Conference (16th, Anchorage, AK, February 7-9, 1990).
PUB TYPE Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Alaska Natives; Bilingual Education; Diagnostic Tests; Elementary Secondary Education; *English (Second Language); Evaluation Criteria; *Language Proficiency; *Language Tests; *Limited English Speaking; *Test Selection
IDENTIFIERS Lower Kuskokwim School District AK

ABSTRACT

In the Lower Kuskokwim School District of western Alaska, the Yupik Eskimo language is the language of preference in most homes and communities, and students are relatively isolated from an environment that requires the consistent use of fluent English. The district's bilingual education staff attempted to locate a commercially developed instrument that could be used to measure the various levels of oral and listening proficiency of their limited English proficient students. The staff agreed that an effective English language proficiency test would: (1) measure a wide range of basic interpersonal communication skills and beginning to intermediate cognitive academic language proficiency skills; (2) assess oral and listening comprehension; (3) provide levels of proficiency, K-12; (4) be diagnostic and prescriptive; (5) be easy to administer, score, and interpret; (6) provide approximate placement to grade level; (7) have a high degree of interrater reliability; and (8) meet testing standards of validity, reliability and appropriate norming. These were the criteria used to evaluate the strengths and weaknesses of the following tests: (1) Language Assessment Scales I and II; (2) Basic Inventory of Natural Language; (3) Idea Oral Language Proficiency Test I and II; (4) Comprehensive English Language Test Secondary; (5) Second Language English Proficiency Test Secondary; and (6) the Maculaitis, K-12. None of the tests was satisfactory. This paper contains recommendations for the limited use of these instruments and suggested features that a more appropriate test would incorporate. (SV)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

English Proficiency Test and Classroom Application

16th Annual Alaska Bilingual Multicultural Education Conference
February 7 - 9, 1990
Anchorage, Alaska

by Bev Williams and Kathy Gross
ESL Teacher Trainers for the Lower Kuskokwim School District

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Bev Williams
Kathy Gross
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED324156

RC017747

In the Lower Kuskokwim School District in the Western part of Alaska, the Yup'ik Eskimo language is spoken as the language of preference in at least 20 of our 23 villages. Many students begin pre-school and kindergarten with little or no comprehension and speaking ability in Standard English. Yup'ik, or a combination of Yup'ik and English provide the major vehicle for communication in the home, in commerce, play, church, and social situations. The village members, separated from the dominant English speaking culture by miles of tundra and limited transportation, rely on the English language in the school environment, but don't rely on it within the confines of the home and community. Our students are relatively isolated from an environment that requires the use of English on a consistent basis with a high degree of fluency. As a result, it appears that most students do not achieve a strong level of proficiency in Standard English even after twelve years of school. Standardized test scores remain low district wide, particulary in the areas of language arts and reading. Students appear to be functioning in English, at least at the communication level, but tend to have difficulty with academic English in oral comprehension and reading. The available information regarding language arts skills and reading ability in English, however, is based on teacher observation and standardized test scores. To date, there is little information available regarding L1 (Yup'ik)

and L2 (English) proficiency. The State of Alaska's Language Assessment system provides information regarding language dominance, but in order to design and implement educational programs in both languages, more specific information is needed regarding language proficiency in L1 and L2. There are still too many unanswered questions without this information.

For example, if the students are dominant in Yup'ik, have they reached a level of proficiency normal for their age group? If the students are dominant in English, are they as proficient as their English speaking peers? If the bilingual or LEP child is having difficulty in school, is it due to limitations of language ability in L1 and/or L2, or is there, a learning disability impairing academic progress that is not solely language based? When do Yup'ik proficient and/or dominant students begin to transfer from Basic Interpersonal Communication Skills (BICS) to more Cognitive Academic Language Proficiency Skills (CALPS)? (Cummins, 1981) Can BICS and CALPS be taught simultaneously?

The selection of an instrument to measure English proficiency, at this point, became our focus. Since Yup'ik is indigenous to this area and is spoken nowhere else in the world, the development of a tool to measure the degree of proficiency in Yup'ik of school age students is the responsibility of the school district and the native communities. For testing English proficiency, however, it was the hope of the LKSD's Curriculum/Bilingual staff, that a commercially developed, validated, reliable and normed instrument could be

located that would measure the various levels and degrees of oral and listening proficiency of the district's K-12 limited English proficient students.

The staff listed the elements that they felt would be essential to collect the type of information necessary for diagnosing student needs, and developing programs and training staff to evaluate, interpret and respond to test results. The staff agreed that an effective English language proficiency test:

1. Measures a wide range of BICs and beginning to intermediate CALPS.
2. Assesses oral and listening comprehension.
3. Provides levels of proficiency, K-12.
4. Is diagnostic and prescriptive.
5. Is easy to administer with minimal training.
6. Is easy to score and interpret.
7. Provides approximate placement to grade level.
8. Has a high degree of interrater reliability.
9. Meets testing standards of validity, reliability and appropriate norming.

1. Measures a wide range of BICS and beginning CALPS. It is generally easy to determine when an individual has very limited ability with a second language, particularly in the beginning stages of basic communication (BICS). It becomes much more difficult to determine oral and listening abilities

when learners gain proficiency with advanced BICS and early CALPS. The problems occur when teachers/testers, unfamiliar with language acquisition theory and with the students' developmental progress, misinterpret casual conversations for fluency. A test is needed to determine at what level of academic proficiency, if any, the student is functioning in relationship to grade level placement.

2. Assesses oral and listening comprehension. There are other instruments currently in use that evaluate reading and writing abilities. Comprehensive Tests of Basic Skills provide one type of information. The District's annual Analytical Writing Assessment provides still another. There are no instruments in place, however, that address the level of listening comprehension of Standard English that occurs in a classroom environment. Nor is there a means for determining if students' communication and academic oral fluency is adequately developed for grade level placement.

3. Provides levels of proficiency, K-12. It would be helpful to teachers/staff developers/bilingual paraprofessionals to have an instrument that would be capable of tracking the language development of students throughout their school history. Teachers would be able to determine language growth and adjust for discrepancies. Staff developers would be able to use the information for curriculum development and selection of resource materials. Bilingual paraprofessionals would have a better understanding of the individual needs of students, and be able to assist Yup'ik speaking students

in their first language within the regular English classroom setting.

4. Diagnostic and prescriptive. For an instrument to be useful to the classroom teachers and curriculum developers, it needs to provide specific information about the individual and group language problems students are encountering. The information should be provided in such a way that would make it possible to prescribe a program/programs for addressing the L2 problems in the classroom and within the curriculum.

5. Is easy to administer with minimal training. When 95% of the student population in a school district are limited English proficient (LEP), and are scattered throughout a geographic area the size of the Republic of South Korea with limited and costly accessibility, language proficiency tests would have to be administered by a classroom teacher with minimal training in second language acquisition and testing. The tests should be group administered, when possible, particulary at the upper elementary and secondary grades. The teachers should be able to administer the tests with minimal or no training, due to not only the lack of ESL experience, but also to the high degree of teacher turnover in rural Alaska. The tests should have a high level of interrater reliability, since many testers/scorers/interpreters would be functioning in dispersed geographic areas.

6. Easy to score and easy to interpret. The same rationale applies to this

component as it does to test administration.

7. Provides approximate placement to grade level. The instrument should provide teachers with realistic analysis of the students' performance in relationship to grade level. Teachers/Students/Parents/Administrators should know if the students' English language oral and listening comprehension will enable them to function and compete academically in a monolingual English speaking setting with confidence and success.

Summary of Tests Reviewed

In examining test review publications (Reviews of English Language Proficiency Tests, edited by Alderson, Drahnke and Stansfield, TESOL, 1987; Large Scale Assessment of Oral Communication Skills: Kindergarten through Grade 12, Edited by Rubin and Mead, ERIC, 1984; Assessment Instruments in Bilingual Education, Center for Bilingual Education, Northwest Regional Education Laboratory, 1978) it should be noted that there are many tests available designed to evaluate English language proficiency. It would be unrealistic to review all instuments available. In reading test reviews, the essential criteria listed above were considered when selecting instruments for piloting. Many instruments were eliminated because they did not address at least 3 or more of the essential components based on the description in reviews. The tests piloted included: LAS I and LAS II (Language Assessment Scales), BINL (Basic Inventory of Natural Language), IPT I and IPT II (Idea Oral Language Proficiency Test), CELT

(Comprehensive English Language Test Secondary), SLEP (Second Language English Proficiency Test-Secondary), and the Maculaitis, K-12.

As the various instruments were evaluated, anecdotal notes were made regarding their strengths and weaknesses in relationship to the desired components.

LAS I & II. The LAS was selected for review because it is developed for students in grades K-12, is designed to measure oral language skills, and theoretically could provide us with information ranging from nonspeakers to fluent speakers of English. It is individually administered, but requires only 20 minutes making it relatively feasible on a large scale. An audio cassette tape is provided, so the oral component can be standardized and consistent throughout the district with different teachers as tests administrators. Also, computerized data management would be available giving a variety of information for analysis.

Strengths> After administering the test to students in grades K-6, with variances in English fluency, several areas of strength were determined. The LAS does an adequate job of testing students new to the English language. It provides an oral retelling of a story, and in doing so, helps assess listening comprehension and oral usage. The listening comprehension component is also effective, since the children when given a sentence are asked to select the picture that most fits the oral discourse. It requires students to understand standard conversational English on a limited scale. It also gives examples of appropriate test item responses for various age levels. The manual provides some enrichment activities. Placement information is available, and very little training is required for test administration.

Weaknesses> The most profound area of weakness, for the purposes of Yup'ik Eskimo students is the fact that the test does not evaluate upper end BICs and beginning CALPS. The child who manages to converse somewhat

fluently, but does not have academic fluency is not identified. Also, the only prescriptive activities suggested are for the LAS's Language Art Series. Training would be required so that teacher administrators could acquire a degree of cross district reliability for scoring the holistic oral component. The secondary section is very similar to the elementary and does not measure academic fluency. The consensus was that, like the elementary, it is designed for the new immigrant child, and not the student whose language is emerging. Little specific diagnostic information is provided, and it must be administered individually, K-12.

BINL. The BINL was examined because it is designed to evaluate oral proficiency in English, can be designed to measure oral proficiency in Yup'ik as well, and can assess language dominance. It also classifies fluency scores by grade level. It can be locally normed, and according to the test's author, Dr. Herbert, the BINL provides both diagnostic and prescriptive information, and can be analyzed by computer program either within the district or by the company. The BINL has been used with other native populations in Alaska, the Inupiaq students of Northwest Arctic and North Slope School Districts.

Strengths> The BINL analyzes sentence structure, oral fluency, vocabulary usage, pronunciation and grammar, as well as listening comprehension. It also examines the natural language of the speaker, and provides the test administrator a recorded oral sample of the student's natural level of fluency.

Weaknesses> On the surface the BINL appeared ideal for this district's purposes, since it can potentially diagnose student oral language and provide a natural sample for before and after comparisons. However, the training requirements for test administrators, the amount of time and training for analysis of data, and the type of data (i.e., metalinguistic in nature) collected proved to be inappropriate for our needs. Because so many different people would be administering the test, maintenance of interrater reliability could be a factor, since results could depend on how well the tester can elicit more language. Although it is said to be prescriptive, it is felt to be limited in the amount of information available. The test also does not measure listening comprehension, an important element, nor does it measure high end BICS, and low end CALPS.

IPT I & II. The Idea Kit and the IPT materials provide an excellent scope and sequence for English language proficiency sequencing, and can be used K-12. It is oral, and evaluates both speaking and listening skills. It also provides placement information for using the Idea Kit which is used in many of the district's schools in the primary and elementary ESL programs. , gives a general grade level placement, and can help teachers diagnose specific language deficiencies.

Strengths> The IPT I and II is easy to administer, requires little training of teachers to administer, score and interpret. It examines student's oral language skills, and identifies some areas of difficulty. It gives prescriptive activities and a starting point within the Idea Kit. It also identifies the students as Fluent, Limited or Non English Speakers (FES, LES, NES) and it is available for use, K-12.

Weaknesses> The IPT is individually administered, and can require up to 30 minutes per student. In order to obtain a high level of interrater reliability, some test administering procedures would need to be taught so that sufficient language is produced for analysis. The IPT does not necessarily measure accurate proficiency levels. Students may miss several items in a lower level, and be able to complete most of the items in a more advanced level. Students can also test out of the kit and still be limited English speakers. Some of the test items may be considered culturally biased for Alaskan Native Students. Also, very little academic language and skills are measured by the test, nor does one know if the students are prepared for mainstreaming in academic settings. When they are considered FES by the IPT, it really means fluency in terms of basic communication, and not English for academic purposes.

CELT (Comprehensive English Language Test). Although it tests secondary only, this test was selected for piloting because it appeared that it could beasily be group administered and scored, and covers several areas of language ability, i.e., listening comprehension, structure, and vocabulary. It also measures intermediate and advanced levels of English for LEP students, which is appropriate for our secondary (9-12) students.

Strengths> The CELT is easily administered, interrater reliability can be maintained with the use of a cassette tape, and is very easy to score. The test also provides information in three important areas of language knowledge for secondary students.

Weaknesses> No diagnostic and prescriptive information is available, unless a lengthy item analysis is conducted to provide this information. There is no measure of oral language proficiency. It is applicable only to grades 9-12, and provides limited placement information. None of the ESL/LEP norming populations used in the CELT approximate the Yup'ik population to be tested. One is left with the overall impression of, "So what?" What good is this information to the classroom teacher, other than knowing what the teacher already knows, ie., the child is not fully proficient in English.

SLEP (Secondary Level English Proficiency). This test is designed for secondary students. It was selected for piloting because it would be easy to administer to a group, easy to score and tests two areas of language ability, listening comprehension and reading comprehension. It places students in six proficiency levels and indicates whether or not the student is ready to be mainstreamed for academic subjects.

Strengths> It is easy to administer and score, and interrater reliability is not a problem due to the cassette tape. The proficiency level descriptions provides useful information regarding potential score expectations on standardized tests in reading and language arts. They also indicates when a student is ready for academic English. The students enjoyed taking this test.

Weaknesses> There is no oral language component. No diagnostic or prescriptive information is provided. Fluent native speakers of English (teachers) taking the test made frequent errors.

**It should be noted that both the CELT and the SLEP have considerable potenial, if an item analysis could be run on items consistently missed, by groups and/or by individual students, making the tests useful for diagnostic as well as placement purposes. The notion of begin able to conduct such an analysis on a wide scale is probably not realistic in terms of time and financial resources.*

Maculaitis. According to the reviews, this test would meet most of the pre-established criteria. It is K-12 in design. It tests listening, speaking, reading and writing. It would select, place, diagnose and determine proficiencies of LEP students. Administration is mostly group, with some individual testing. With the answer and score sheets, it appears easy to administer and score. The reviews contend that it has high reliability.

Strengths> The test does cover the four language arts areas as well as grades K-12. It is relatively quick to score, and is group administered in grades 4-12, except for the oral component. At the secondary level, two types of writing are assessed, a narration and an application form. The administration is simple since the examiner's manual is self-explanatory. The cassette tape for the listening portion helps to maintain interrater reliability.

Weaknesses> It took over 2 hours to administer all portions of the test to one group not counting the individual oral portion. This is entirely too lengthy and cumbersome for district wide testing, considering the other standardized tests currently mandated and administered. The interpretation of scores is too cumbersome to be used at each village site. Training would be required for holistic scoring in the written and oral portion, and maintaining interrater reliability might become a crucial factor. It did not provide diagnostic information as had been expected.

Conclusions

After an intensive search for a test designed to evaluate students' levels of English proficiency for measuring academic success, no satisfactory test was found. Classroom teachers need more than just the statistics that most of these tests provide. They need prescriptive and diagnostic information, beneficial to the development of lessons, units and curriculum for the limited English proficient students.

As potential users of these tests, we were basically disappointed in what we

found available for our population of LEP students who are not new immigrants and who cover the complete range of beginning BICS to advanced CALPS. The tests were either concentrating on beginning skills totally, or tested everything in depth using the same format as standardized basic skills tests, collecting much of the same information. As stated by Alba Ortiz, Director of the Office of Bilingual Education at The University of Texas at Austin, "...For the most part, state approved language proficiency instruments test BICS. Consequently, there is a danger that children will be prematurely exited from special language programs." (p. 2) Or as in our case, be treated as though there are no language problems, or if there are, the problems are regarded as learning disabilities (communication disorders) as opposed problems dealing with limited English proficiencies.

Most tests did not provide the kinds of diagnostic/prescriptive information useful to the classroom teacher and curriculum developers. We want to know more than whether or not our students are LEP. That we know. We also want to know whether or not they are beginning, middle, or advanced limited English proficient (LEP) in BICS and/or CALPS.

Most of the tests provided us with Non English, Limited English, and Fluent English Proficient information, but the term fluency is not referring to academic fluency and when results are given and this term fluency is used, it is misleading and misinterpreted. It certainly does not refer to the same level of fluency as defined by the Foreign Service Institute Oral Proficiency

Interview, and the American Council on the Teaching of Foreign Language/Educational Testing Service Proficiency Guidelines. With these guidelines a person nearing fluency, or scoring a 4, has English language skills in all topics "...normally pertinent to professional needs and is nearly equivalent to ENS (English Native Speaker)." A person achieving fluency and scoring a 5, can function at an equivalent level to an educated native speaker on all subjects. (Reviews of English Language Proficiency Tests, p. 45) Our LEP students do not achieve either level, for the most part, by twelfth grade. In terms of the information provided in most proficiency tests, the students are viewed as having far more English language proficiency than truly exists, at least for classroom purposes. We need to know what their ability/fluency/proficiency is in English in relationship to what is required in content area classes at the various grade levels, and we would like to be able to monitor their progress as they work at attaining fluency.

Most teachers can identify BICS with some training. The difficulties occur when students sound fluent, but aren't comprehending what they hear or read, and cannot express themselves coherently in writing or in oral discourse when dealing with classroom information.

We are still searching for the test that will measure NES, LES and FES in relationship to academic language. If one exists, and our search is indeed incomplete, we need to know. If one does not exist, this could be

considered a message to testing companies who will find that there are many schools in similar situations, educators who are finding that they cannot get adequate information regarding LEP students from traditional assessment tools and can find nothing on the market designed for bilingual students that adequately evaluate the students' proficiencies. As is stated in the Assessment of Language Minority Students: A Handbook for Educators, Illinois Resource Center (1985), "Although...tools such as standardized, commercially developed English proficiency tests exist, their use with LEP students has serious drawbacks. The students' lack of facility with English impedes their performance, making it difficult to obtain an accurate assessment of skills. Also, because these standardized tests measure only select aspects of language, they do not reflect these students' overall proficiency and comprehension." (p. 14)

Recommendations for the Lower Kuskokwim School District

Immediate Action-Primary Elementary: We recommend the IPT for grades K-8 for all students who have been identified as LEP in the Alaska State Language Assessment and/or who are scoring one or more grades below grade level in reading and language arts on the district's selected standardized test. The IPT follows a scope and sequence and grade level, and it can provide diagnostic information with relative ease so that teachers have information in hand to deal with language difficulties.

ESL outcomes have been developed and reviewed by ESL teachers for grades

K-6, and will be piloted in our district during the 1990-91 school year. We recommend that these be used as entry/exit criteria for grade levels, so that children aren't continually set up for failure by being promoted to grade levels requiring higher proficiencies of English when such proficiencies do not exist. The outcomes provide diagnostic information, and teachers can, with some training, *prescribe* activities that address the needs.

Immediate Action-Secondary: Neither of the two tests that we would like to consider using provides sufficient information to warrant an unconditional recommendation. The CELT tests more areas, but provides little diagnostic information without running an extensive item analysis. The SLEP provides more useful information, but only examines listening and reading. We would like to continue using both of these tests in the fall of 1990 before an overall district recommendation is made.

Continued use of the Analytical Writing Assessment, which we've been conducting for the past three years, will provide information on LEP students' ability to use English in relationship to their monolingual English speaking peers. We will determine on a yearly basis if the gap is narrowing or not.

Long Range Action: We realize that the conditions for testing that we established early on may not be realistic in terms of finding a commercial test that would meet all of the criteria. After examining the Assessment of

Language Minority Students (cited earlier) we are beginning to feel that in order to obtain the type of information that will be most useful to teachers and staff developers, a test or tests will need to be designed to do this. The test could contain the following features:

1) A cloze test using passages from content area texts to help determine the grade level and varying levels of ability at which students are functioning academically with reading comprehension. Further oral and written cloze tests could examine grammar, word usage and vocabulary knowledge in communicative and academic situations.

2) A dictation test that would measure holistically the student's ability "...to reproduce in writing a passage that is heard orally." (Assessment of Language Minority Students. Illinois Resource Center, 1985). The right type of dictation could help determine how well the student is able to process information at various levels of language ability. (Cohen, A. 1989.*Testing Language Ability in the Classroom.)*

3) A Listening Comprehension test to determine the degree of comprehension occurring in a content area class at all grade levels, could be evaluated with the listening portion of the commercially developed tests, such as the Maculaitis, the SLEP or CELT. (Schools in Oregon have expressed interest in determing listening skills of students. Interwest Inc. is currently working on developing a Listening Comprehension rubric very

much like the Analytical Writing Assessment rubric to be used with all students. The development of such an instrument would be quite useful in our district.)

5) Continued use of the Analytical Writing Assessment currently in use in the district . "Obtaining a short writing sample and rating it in a holistic integrative manner is very similar to obtaining and rating an oral language sample..." (Assessment of Language Minority Students, p. 11) As stated by Muriel Saville Troike in the article *Teaching and Testing Both Communicative and Academic skills in English as a Second Language,* (Issues of Language Assessment 1983, p. 141), the language skill most likely to develop competence in using and understanding language in context -reduced situations (which is what most classrooms are like) is writing. Analysis of ESL writing should help us determine the degree to which students are prepared for mainstream classroom work.

6) Criterion-reference tests need to be continually developed and improved in all content areas. Time and resources should be made available so that student outcomes for each subject at each grade level can be determined by teachers. Then, realistic and uniform standards can be used to determine if students do indeed meet entry criteria and can exit to the next grade, taking with them the skills, knowledge and language necessary to express conceptual understanding for academic success at the next level.

Further Recommendations:

Limited English proficient students should not be tested with standardized tests before grade 4. Their reading/thinking/conceptual development in English is not sufficient for the tests to measure any knowledge or skills. The Yup'ik CRTs are useful to the classroom teacher to help diagnose and prescribe for an individual child. The CTBS/ITBS only tell us what we already know....the students are not ready to function academically in English.